

# A modified of nonmonotone spectral conjugate gradient method for unconstrained optimization

**ZiXing Rong**

College of Mathematics and Computer Science, Hebei University, Baoding, China

**Ke Su**

College of Mathematics and Computer Science, Hebei University, Baoding, China

**Bei Gao**

College of Mathematics and Computer Science, Hebei University, Baoding, China

## Abstract

According to a modified parameter  $\beta_k$ , a new nonmonotone spectral Conjugate Gradient method for unconstrained optimization is proposed in this paper, which combines the conjugate gradient direction with spectral step-length effectively. We applied the spectral step-length to the entire Conjugate Gradient direction rather than the negative gradient direction, and take advantage of the new nonmonotone F-rule for line searches to obtain the next iteration point. The global convergent property of the algorithm with the modified parameter and the proposed nonmonotone F-rule for line search are proved under some appropriate conditions.

**Mathematics Subject Classification:** Unconstrained optimization

**Keywords:** unconstrained optimization, spectral conjugate gradient method, global convergence, nonmonotone line search

## 1 Introduction

Consider the following unconstrained optimization problem

$$\min f(x), x \in R^n \quad (1)$$

where  $f : R^n \rightarrow R$  is a continuously differentiable function,  $R^n$  is an Euclidean space.

Unconstrained optimization problem is an important research topic in mathematical programming fields. There are some methods for solving unconstrained optimization problem, such as feasible direction algorithm, gradient-type methods, Newton-type and so on. It is well-known that conjugate gradient

method is a good method for solving the unconstrained optimization problem in management and engineering. It has the following formula for solving problem (1):

$$x_{k+1} = x_k + \alpha_k d_k$$

$$d_k = \begin{cases} -g_k, & k = 0, \\ -g_k + \beta_k d_{k-1}, & k \geq 1. \end{cases}$$

where  $d_k$  is a search direction,  $\beta_k$  is a scalar,  $g_k = \nabla f(x_k)$ ,  $\alpha_k$  is a positive step-size along the search direction.

In many literatures (refer to [1-3]), the main thing is to choose the scalar  $\beta_k$ , which leads to many different conjugate gradient methods according to the different  $\beta_k$ . A well-known conjugate gradient method was proposed by Fletcher and Reeves [1], and some other choices as followed:

$$\beta_k^{HS} = \frac{g_k^T y_{k-1}}{d_{k-1}^T y_{k-1}}, \quad \beta_k^{FR} = \frac{\|g_k\|^2}{\|g_{k-1}\|^2},$$

$$\beta_k^{PRP} = \frac{g_k^T y_{k-1}}{\|g_{k-1}\|^2}, \quad \beta_k^{CD} = -\frac{\|g_k\|^2}{d_{k-1}^T g_{k-1}},$$

$$\beta_k^{LS} = -\frac{g_k^T y_{k-1}}{d_{k-1}^T y_{k-1}}, \quad \beta_k^{DY} = \frac{\|g_k\|^2}{d_{k-1}^T y_{k-1}},$$

where  $\|\cdot\|$  denotes Euclidean norm,  $y_{k-1} = g_k - g_{k-1}$ .

Raydan introduced the spectral gradient method (SGM) for potentially large-scale unconstrained optimization [4]. The main feature of this method is that only gradient directions are used at each line search whereas a nonmonotone strategy guarantees global convergence. The idea of nonmonotone technique can be traced back to Grippo et al. in 1988 [5], thanks its excellent numerical exhibition, many nonmonotone techniques have been developed in recent years, for example, nonmonotone line search approaches, and nonmonotone trust region methods.

Motivated by the ideas of spectral gradient and conjugate gradient methods [6-9], we combined the new nonmonotone techniques with the spectral gradient method to obtain a more efficient algorithm. Different from the existing research results, the paper will propose a new class of nonmonotone spectral conjugate gradient method. the paper will modify the spectral gradient method of [10] from the parameter  $\beta_k$ , in addition, We will apply spectral step-length to the entire Conjugate Gradient direction instead of the negative gradient direction, and take advantage of the new nonmonotone F-rule proposed by Yu ZhenSheng for line searches to obtain the next iteration point.

The paper is organized as follows. in section 2, The modified algorithm is proposed, all the essential features of its implementation is given, in section 3, we establish the global convergence of the algorithm.

## 2 A modified of nonmonotone spectral conjugate gradient method

Assume that  $f : R^n \rightarrow R$  has continuous partial derivatives. The problem considered in this paper is

$$\min f(x), x \in R^n \quad (2)$$

The iteration formulation for (2) is as follows:

$$x_{k+1} = x_k + \alpha_k d_k,$$

where  $d_k$  is a search direction and  $\alpha_k$  is a step length chosen to induce the value of  $f(x)$  The direction is generated by

$$d_k = -\theta_k g_k + \beta_k d_k,$$

where  $g_k$  denotes  $\nabla f(x_k)$  and  $d_0 = -\theta_0 g_0$  and  $\theta_k$  is a arbitrary parameter

$$\beta_k = \frac{(\theta_k y_{k-1} + s_{k-1})^T g_k}{d_{k-1}^T y_{k-1}},$$

where  $s_k = x_{k+1} - x_k = \alpha_k d_k$  and  $y_k = g_{k+1} - g_k$ .

In general, we denote  $\theta_k$  as

$$\theta_k = \frac{s_{k-1}^T s_{k-1}}{s_{k-1}^T g_{k-1}} ([10])$$

It's easy to prove that, the spectral conjugate gradient method does not guarantee  $d_k$  to be the decent direction, which leads to the difficulties on the global convergence. To tackle this problem, many authors proposed various methods, including nonmonotone technique.

To describe the nonmonotone technique, we describe the nonmonotone Armijo rule.  $\alpha_k$  is a stepsize with  $\alpha_k \geq 0$  and  $d_k$  is a search direction satisfied  $g_k^T d_k \leq 0$ , Let  $a > 0$ ,  $\gamma \in (0, 1)$ ,  $\beta \in (0, 1)$  and let  $M$  be a nonnegative integer. For each  $k$ , let  $m(k)$  satisfies

$$m(0) = 0, 0 \leq m(k) \leq \min[m(k-1) + 1, M], \text{ for } k \geq 1,$$

Let  $\alpha_k = \beta^{p^k} a$  and  $p^k$  be the smallest nonnegative integer  $p$  such that

$$f(x_k + \alpha_k) \leq \max_{0 \leq j \leq m(k)} [f(x_{k-j})] + \gamma \beta^{p^k} a g_k^T d_k.$$

If the search is the nonmonotone Goldstein line search, so  $\alpha_k$  should satisfied the following condition:

$$f(x_k + \alpha_k) \leq \max_{0 \leq j \leq m(k)} [f(x_{k-j})] + \mu_1 \lambda_k g_k^T d_k,$$

$$f(x_k + \alpha_k) \leq \max_{0 \leq j \leq m(k)} [f(x_{k-j})] + \mu_2 \lambda_k g_k^T d_k,$$

where  $0 < \mu_1 \leq \mu_2 < 1$

If the search is the nonmonotone wolfe line search, so  $\alpha_k$  should satisfied the following condition:

$$f(x_k + \alpha_k) \leq \max_{0 \leq j \leq m(k)} [f(x_{k-j})] + \gamma_1 \alpha_k g_k^T d_k,$$

$$g(x_k + \alpha_k d_k)^T d_k \geq \gamma_2 g_k^T d_k,$$

where  $0 < \gamma_1 \leq \gamma_2 < 1$

Now we give a new nonmonotone F-rule line search proposed by Yu Zhen-Sheng as follows: [11]

Let  $\lambda \in (0, 1]$   $M \geq 1$  is a positive integer, defined  $m(k) = \min[k + 1, M]$

$$\lambda_{kr} \geq \lambda, \quad r = 0, 1, 2, \dots, m(k) - 1 \quad \sum_{r=0}^{m(k)-1} \lambda_{kr} = 1,$$

Let  $\alpha_k \geq 0$  be bounded above and satisfy:

$$f(x_k + \alpha_k d_k) \leq \max[f(x_k),$$

$$\sum_{r=0}^{m(k)-1} \lambda_{kr} f(x_{k-r})] + \gamma \alpha_k \langle d_k, g(x_k) \rangle.$$

We now present the modified algorithm of the nonmonotone spectral conjugate gradient method. In this method, we give the conjugate gradient direction, where the parameter  $\beta_k$  is chosen from *Yasushi Narushima* and *Hiroshi Yabe* [12] and then obtain the a convex combination of conjugate direction, which enhances the convergent properties.

Set  $d_k^1$

$$d_k^1 = -\mu g_k + (1 - \mu) \beta_k d_{k-1},$$

where  $\mu \in (\frac{1}{2}, 1]$  and

$$\beta_k^{YH} = \begin{cases} 0, & \text{if } g_k^T d_{k-1} \leq 0, \\ \frac{\|g_k\|^2}{(g_k^T d_{k-1} + \|g_k^T\| \|d_{k-1}\|)}, & \text{otherwise.} \end{cases}$$

and, we define spectral conjugate gradient as follow:

$$d_k = \theta_k d_k^1. \quad (3)$$

The definition of parameter  $\theta_k$  will be given in the following algorithm.

**Algorithm 2.1**

**Step0.** Give a positive integer  $M$  and some constants,  $\varepsilon > 0$ ,  $1 > \sigma_2 > \sigma_1 > 0$ ,  $\gamma \in (0, 1)$ ,  $\delta > 0$ ,  $\theta_{\max} > \theta_{\min} > 0$ , set  $\theta_0 \in [\theta_{\max}, \theta_{\min}]$ ,  $k = 0$ .

**Step1.** Compute  $g_k$ , and  $\|g_k\| \leq \varepsilon$ , STOP.

**Step2.** Compute  $d_k$  by (3).

**Step3.1** Set  $\alpha_k \leftarrow \frac{-\delta g_k^T d_k}{\|d_k\|^2}$ .

**Step3.2** Set  $x_+ = x_k + \alpha_k d_k$ .

**Step3.3** Let  $\lambda \in (0, 1]$ , define  $m(k) = \min[k + 1, M]$  Let

$$\lambda_{kr} \geq \lambda, r = 0, 1, 2, \dots, m(k) - 1 \quad \sum_{r=0}^{m(k)-1} \lambda_{kr} = 1.$$

if

$$f(x_k + \alpha_k d_k) \leq \max\{f(x_k), \sum_{r=0}^{m(k)-1} \lambda_{kr} f(x_{k-r})\} + \gamma \alpha_k \langle d_k, g(x_k) \rangle \quad (4)$$

define  $x_{k+1} = x_+$ ,  $s_k = x_{k+1} - x_k$ ,  $y_k = g(x_{k+1}) - g(x_k)$ , and go to step 4.

if (3) does not hold, define  $\alpha_{new} \in [\sigma_1 \alpha_k, \sigma_2 \alpha_k]$  let  $\alpha_k = \alpha_{new}$  and go to step (3.2).

**Step4.** Compute  $b_k = \langle s_k, y_k \rangle$ , if  $b_k \leq 0$ , set  $\theta_{k+1} = \theta_{\max}$ , otherwise compute  $a_k = \langle s_k, s_k \rangle$  and

$$\theta_{k+1} = \min\{\theta_{\max}, \max\{\theta_{\min}, a_k/b_k\}\}.$$

**Step5.** Set  $k := k + 1$ , and go to step 1 .

### 3 The global convergence of the algorithm

In order to achieve the convergence of the algorithm ,we give some Assumptions and Lemmas as follow:

**Assumption 3.1**

A:  $f(x)$  is bounded above on the level set  $L = \{x | f(x) \leq f(x_0)\}$

B: In some neighborhood  $\Omega$  of  $L$  , $f$  is continuously differentiable, and its gradient  $g$  is Lipschitz continuous, namely, there exists a constant  $L$  such that

$$\|g(x) - g(y)\| \leq L\|x - y\|$$

**Lemma 3.1** Assume  $d_k$  is generated by the algorithm (2.1), then

$$g_k^T d_k \leq \theta_{min}(1 - 2\tau)\|g_k\|^2. \quad (5)$$

where  $\tau \in (\frac{1}{2}, 1)$

Proof:

if  $\beta_k = 0$ , then  $d_k = -\theta_k \mu g_k$ , since,  $\theta_k \geq \theta_{min}$  and  $\mu \in (\frac{1}{2}, 1]$ , there exist  $\tau \in (\frac{1}{2}, 1]$ , and  $\mu > \tau$ , so we have

$$g_k^T d_k = -\theta_k \mu \|g_k\|^2 \leq -\theta_{min} \tau \|g_k\|^2 \leq \theta_{min}(1 - 2\tau)\|g_k\|^2$$

if  $\beta_k > 0$ , by the definition of  $d_k$ , we have

$$\begin{aligned} g_k^T d_k &= \theta_k(-\mu \|g_k\|^2 + (1 - \mu)\beta_k g_k^T d_{k-1}) \\ &\leq \theta_k(-\mu \|g_k\|^2 + (1 - \mu)\frac{\|g_k\|^2}{2g_k^T d_{k-1}} g_k^T d_{k-1}) \\ &\leq \theta_k(-\tau \|g_k\|^2 + (1 - \tau)\|g_k\|^2) \\ &\leq \theta_k(1 - 2\tau)\|g_k\|^2 \\ &\leq \theta_{min}(1 - 2\tau)\|g_k\|^2. \end{aligned}$$

so (5) hold.

**Lemma 3.2** Assumption  $B$  is hold,  $\alpha_k$  satisfy the formula (4) of algorithm, then there exist  $\pi_k \in [\sigma_1, \sigma_2]$  satisfy :

$$\alpha_k \geq \min\left\{1, \frac{(1 - \gamma)\pi_k |\langle g_k, d_k \rangle|}{L \|d_k\|^2}\right\}. \quad (6)$$

Proof:

At the  $k$ th iterate, if  $\beta = 1$  satisfy the formula (4). then  $\alpha_k = 1$ ,

otherwise, there exist  $\pi_k \in (\sigma_1, \sigma_2)$ , which does not satisfy formula (4) for  $\beta_k/\pi_k > 0$ , in other words :

$$\begin{aligned} f(x_k + \frac{\alpha_k}{\pi_k} d_k) &> \max\{f(x_k), \sum_{r=0}^{m_k-1} \lambda_{kr} f(x_{k-r})\} + \gamma \frac{\alpha_k}{\pi_k} \langle d_k, g(x_k) \rangle \\ &> f(x_k) + \gamma \frac{\alpha_k}{\pi_k} \langle d_k, g(x_k) \rangle. \end{aligned} \quad (7)$$

by mean value theorems, we have:

$$\begin{aligned} f(x_k + \alpha d_k) - f(x_k) &= \int_0^\beta \langle g(x_k + td_k) - g(x_k), d_k \rangle dt + \alpha \langle g(x_k), d_k \rangle \\ &\leq \frac{1}{2} L \alpha^2 \|d_k\|^2 + \alpha \langle g(x_k), d_k \rangle. \end{aligned}$$

with (7) , we have

$$\alpha_k \geq \min\left\{1, \frac{(1-\gamma)\pi_k}{L} \frac{|\langle g_k, d_k \rangle|}{\|d_k\|^2}\right\}.$$

So, (6) is hold.

**Lemma 3.3** Assume that sequence  $x_k$  is generated by algorithm 2.1, so that:

$$\begin{aligned} f(x_k) &\leq f(x_0) + \lambda\gamma \sum_{r=0}^{k-2} \alpha_r \langle g(x_r), d_r \rangle + \gamma\alpha_{k-1} \langle g(x_{k-1}), d_{k-1} \rangle \\ &\leq f(x_0) + \lambda\gamma \sum_{r=0}^{k-1} \alpha_r \langle g(x_r), d_r \rangle. \end{aligned} \tag{8}$$

Proof: We prove by induction.

if  $k = 1$ , by (4) and  $\lambda \leq 1$  , we have

$$f(x_1) \leq f(x_0) + \lambda\alpha_0 \langle g(x_0), d_0 \rangle \leq f(x_0) + \gamma\lambda\alpha_0 \langle g(x_0), d_0 \rangle.$$

Assume (8) is hold for  $1, 2, \dots, k$  , we can think of this problem from two cases :

case 1 :  $\max[f(x_k), \sum_{r=0}^{m_k-1} \lambda_{kr} f(x_{k-r})] = f(x_k)$ , by (4),we have

$$\begin{aligned} f(x_{k+1}) &= f(x_k + \alpha_k d_k) \leq f(x_k) + \gamma\alpha_k \langle g(x_k), d_k \rangle \\ &\leq f(x_0) + \lambda\gamma \sum_{r=0}^{k-1} \alpha_r \langle g(x_r), d_r \rangle + \gamma\alpha_k \langle g(x_k), d_k \rangle \\ &\leq f(x_0) + \lambda\gamma \sum_{r=0}^k \alpha_r \langle g(x_r), d_r \rangle. \end{aligned}$$

case 2 :  $\max[f(x_k), \sum_{r=0}^{m_k-1} \lambda_{kr} f(x_{k-r})] = \sum_{r=0}^{m_k-1} \lambda_{kr} f(x_{k-r})$ , Let  $q = \min[k, M-1]$  by (4) we have

$$\begin{aligned} f(x_{k+1}) &= f(x_k + \alpha_k d_k) \\ &\leq \sum_{p=0}^q \lambda_{kp} f(x_{k-p}) + \gamma\alpha_k \langle g(x_k), d_k \rangle \\ &\leq \sum_{p=0}^q \lambda_{kp} [f(x_0) + \lambda\gamma \sum_{r=0}^{k-p-2} \alpha_r \langle g(x_r), d_r \rangle + \gamma\alpha_{k-p-1} \langle g(x_{k-p-1}), d_{k-p-1} \rangle] \\ &\quad + \gamma\alpha_k \langle g(x_k), d_k \rangle. \end{aligned}$$

impose  $(1, 2, \dots, q) \times (1, 2, \dots, k-q-2) \subset \{(p, r) : 0 \leq p \leq q, 0 \leq r \leq k-q-2\}$ ,  
 $\sum_{p=0}^q \lambda_{kp} = 1$ ,  $\lambda_{kp} \geq \lambda$ , we have

$$\begin{aligned}
f(x_{k+1}) &\leq f(x_0) + \lambda \sum_{r=0}^{k-q-2} \left( \sum_{p=0}^q \lambda_{kp} \right) \alpha_r \langle g(x_r), d_r \rangle + \gamma \sum_{p=0}^q \lambda_{kp} \alpha_{k-p-1} \\
&\quad \langle g(x_{k-p-1}), d_{k-p-1} \rangle + \gamma \alpha_k \langle g(x_k), d_k \rangle \\
&\leq f(x_0) + \lambda \gamma \sum_{r=0}^{k-q-2} \alpha_r \langle g(x_r), d_r \rangle + \lambda \gamma \sum_{r=k-p-1}^{k-1} \alpha_r \langle g(x_r), d_r \rangle \\
&\quad + \gamma \alpha_k \langle g(x_k), d_k \rangle \\
&= f(x_0) + \lambda \gamma \sum_{r=0}^{k-1} \alpha_r \langle g(x_r), d_r \rangle + \gamma \alpha_k \langle g(x_k), d_k \rangle \\
&\leq f(x_0) + \lambda \gamma \sum_{r=0}^k \alpha_r \langle g(x_r), d_r \rangle
\end{aligned}$$

end.

**Theorem 3.1** Assume that  $x_k$  and  $d_k$  is generated by algorithm 2.1 and  $A$ ,  $B$  holds, and then

$$\lim_{k \rightarrow \infty} \langle g(x_k), d_k \rangle = 0. \quad (9)$$

Proof:

Assume that there exist a boundless sequence index set  $K$ , and there exist  $\varepsilon > 0$ , which satisfy  $\langle g(x_k), d_k \rangle \leq -\varepsilon$ . for any  $k \in K$ , impose the lemma 3.3, for any  $k \in K$ , we have :

$$-\lambda \gamma \sum_{r=0}^{k-1} \alpha_r \langle g(x_r), d_r \rangle \leq f(x_0) - f(x_k), \quad (10)$$

impose the lemma 3.1,  $g_k^T d_k \leq \theta_{\min}(1 - 2\tau) \|g_k\|^2$ , which means

$$-\frac{\langle g(x_k), d_k \rangle}{\|g_k\|^2} \geq \theta_{\min}(2\tau - 1), \quad (11)$$



with (6) (10), (11) we have:

$$\begin{aligned}
f(x_0) - f(x_k) &\geq -\lambda\gamma \sum_{r=0, r \in K}^{k-1} \alpha_r \langle g(x_r), d_r \rangle \\
&\geq \lambda\gamma\varepsilon \sum_{r=0, r \in K}^{k-1} \alpha_r \\
&\geq \lambda\gamma\varepsilon \sum_{r=0, r \in K}^{k-1} \min\left\{1, \frac{(1-\gamma)\pi_r}{L} \cdot \frac{|\langle g(x_r), d_r \rangle|}{\|d_r\|^2}\right\} \\
&\geq \lambda\gamma\varepsilon \sum_{r=0, r \in K}^{k-1} \min\left\{1, \frac{(1-\gamma)\sigma_1}{L} \cdot \theta_{\min}(2\tau - 1)\right\}
\end{aligned}$$

since  $f(x)$  is bounded below, let  $k \rightarrow \infty (k \in K)$ , we have

$$\infty \geq f(x_0) - f(x_k) \rightarrow \infty.$$

paradoxically, so (9) holds  
end.

**Theorem 3.2** Assume that sequence  $\{x_k\}$  is generated by algorithm 2.1, and then we have

$$\lim_{k \rightarrow \infty} \|g_k\| = 0$$

Proof:

by Lemma 3.1 and definition 3.1, we have

$$0 \geq \lim_{k \rightarrow \infty} -\frac{\theta_{\min}}{2} \|g_k\|^2 \geq \lim_{k \rightarrow \infty} \langle g(x_k), d_k \rangle = 0$$

which implies

$$\lim_{k \rightarrow \infty} \|g_k\| = 0$$

end.

## References

- [1] Fletcher, Reeves, C. M, Function Minimization by Conjugate gradients, Computer Journal, 1964, Vol7, 149-154.
- [2] Fletcher, R, Practical Methods of Optimization, Wiley, New York, 1987, Vol I, Unconstrained Optimization, 2nd edition.
- [3] Dai, Y.H. and Yuan, Y, A nonlinear conjugate gradient methods with a strong global convergence property, SIAM J Optim. 10, 1999, 177-182.

- [4] Raydan M, The Barzilai and Borwein gradient method for the large unconstrained minimization problem, *SIAM Journal on Optimization* 7, 1997, 26-33.
- [5] L.Grippo,F.Lampariello,S.Ludidi, A nonmonotone line search technique for Newton's method, *SIAM J.Numer.Anal.* 23, 1986, 707-716.
- [6] Birgin E. G, Chambouleyron I. and Martinez J. M, Estimation of optical constants of thin films using unconstrained optimization, *Journal of Computational Physics* 151, 1999, 862-880.
- [7] Glunt W. Hayden T. L. and Raydan M, Molecular conformations from distance matrices, *Journal of Computational Chemistry* 14, 1993, 114-120.
- [8] Luengo F. Raydan M, Glunt W. and Hayden T. L, Preconditioned spectral gradient method for unconstrained optimization problems, *Escuela de Computacin, Facultad de Ciencias,Universidad Central de Venezuela, 47002 Caracas 1041-A,Venezuela, 1996, Technical Report 96-08.*
- [9] Raydan M, The Barzilai and Borwein gradient method for the large unconstrained minimization problem, *SIAM Journal on Optimization* 7, 1997, 26-33.
- [10] Birgin,E. and Martinez,J.M, A spectral conjugate gradient method for unconstrained optimization, *Applied Mathematics and Optimization.* 43, 2001, 117-128.
- [11] Yu Z S,Pu D G, A new nonmonotone line search technique for unconstrained optimization, *Journal of Computational and Applied Mathematic,* 219, 2008, 134-144.
- [12] Narushima,Y. and Yabe,H, Global convergence of a memory gradient method for unconstrained optimization, *Computation Optimization and Application,* 35, 2006, 325-346.

**Received: January, 2015**